Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

AI Agency Risks and their Mitigation through Business Process Management: a Conceptual Framework

Anna Sidorova
University of North Texas
Anna.sidorova@unt.edu

Dana Rafiee
Destiny Corporation
drafiee@destinycorp.com

Abstract

After over 60 years of research and development, AI has made its way into mainstream business operations. Continuous advances in the fields of machine learning, knowledge representation, and logical reasoning are expected to result in higher autonomy of AI-enabled systems such as Distributed AI (DAI) agents that can think and act. The increased agency of the AI systems is expected to result in agency risks and the need for mitigating such risks through AI governance. In this paper, we build on agency theory and identify factors that increase the risk of an agency problem between a principal (a human or an organization) and an AI agent and propose a framework for AI agency problem analysis. The framework is illustrated through AI use cases and industry examples. Implications for AI governance research and practice are discussed.

1 Introduction

After over 60 years of research and development, AI has made its way into mainstream business operations as well as the personal life of unsuspecting individuals. According to the McKinsey Global Institute, "it is poised to cause the next wave of digital disruption" [1, p. 6] and Gartner anticipates that by 2022, AI will know more about the emotional state of an individual than the people they are closest to. Over 60% of personal device vendors will rely on third-party cloud AI services [2]. At the enterprise level, innovations such as RAGE-AI promise "zero-code, model-driven software development using highly abstract components, and traceable machine learning" [3, p. 5]. As more activities are transferred from human actors and code-driven IT to model driven and ML based solutions, organizations need to devise new types of control mechanisms to ensure that the goals of the AI artifacts are aligned with those of organizational stakeholders. Gartner makes another prediction, this time stating there will be a rise in the percentage of workers dedicated to monitoring and guiding neural networks, a popular class of machine learning algorithms [3]. Most commercial applications of AI today are relatively narrow in scope. Image recognition, natural language processing, predictive models for geological exploration, and generative models used for automatic translation are highly dependent on their human handlers for data and process input, along with model tuning. However, AI artifacts are expected to become increasingly autonomous, posing the risk of an agency problem for users. Increased autonomy of AI-enabled artifacts calls for the development of AI governance frameworks that would help in the establishment of policies concerning the development of AI, as well as the actions of AI agents, and guide the monitoring of the proper implementation of such policies.

Business Process Management is a diverse research field that emerged at the inter-section of three process management traditions, the quality control/scientific management tradition, the strategic management tradition, and the information technology tradition [4]. The managerial tradition within BPM research is concerned with the alignment be-tween strategic goals of an organization and its business processes [4]. As managerial BPM research is directly concerned with organizational governance issues [10], it is uniquely positioned to tackle the issue of AI governance. However, our understanding of how BPM can help better manage the opportunities and challenges presented by the introduction of advanced analytics and AI into business operations is still in its nascent form. We seek to identify ways in which BPM can contribute to AI governance by adopting the agency theory perspective [11]. Building on the agency theory, we identify factors that increase the risk of the agency problem in the organizational use of AI and develop a framework for the AI agency problem analysis. Guided by such a framework, we seek to identify AI governance policies can contribute to reducing or mitigating AI agency risks.

URI: https://hdl.handle.net/10125/60019
ISBN: 978-0-9981331-2-6

Specifically, we seek to answer the following research questions:

1. What factors increase the risk of the agency risks in organizational use of AI?
2. How can BPM help address the AI agency risks?

In the next section, we provide a brief background on AI, also known as Machine Intelligence, as well as the key tenets of the agency theory. We then develop a conceptual framework for analyzing agency risks associated with the organizational use of AI and illustrate it using industry examples and AI use case vignettes. Next, we examine how various governance policies can be helpful in addressing the AI agency problem and identify roadblocks that need to be addressed. Finally, implications for AI governance researchers and practitioners are discussed.

2 Theory and background

2.1 The rise of Artificial Intelligence

The term artificial intelligence was coined in 1955 and the Turing test for artificial intelligence was famously described in a 1950 publication [4], [5]. Since then, the AI field has endured periods of fast progress punctuated with periods of relative inactivity as the AI community has not been able to meet the inflated expectations of its stakeholders [6]. In the early 2000s, a confluence of several technological trends led to an exponential growth in AI capabilities. Several important milestones followed: the development of a self-driving car, increased accuracy of the identification of fraud in financial transactions, and the development of AI algorithms capable of beating human contestants in such games as Jeopardy, Go, and Poker [7, p. 9]. While most currently available AI systems are narrow in scope and designed to excel in performing specific tasks, significant progress in areas such as representation learning, transfer learning, and reinforcement learning is contributing to the development of Artificial General Intelligence. For example, while facial recognition is available today, the Deep Learning algorithms can determine anyone's mood based on their facial expressions and actions.

2.2 Agency theory

Agency theory was developed as a means for examining situations in which cooperating parties have different goals or attitudes towards work or risk [8]. It has since been applied to executive and employee compensation, inter-firm contract design, and more. [9]–[12]. At the core of the theory lies the relationship between a principal (a party who delegates the work) and an agent (a party who performs the work). Most applications of agency theory are focused on resolving the agency problem, which can arise "when (a) the desires or goals of the principal and agent conflict and (b) it is difficult or expensive for the principal to verify what the agent is actually doing" [13, p. 58].

Governance can be generally defined as a set of structures and processes put in place to mitigate the agency problem within an organization or a process, e.g. to ensure that actors performing specific activities do so in such a way as to maximize goal attainment of organizational stakeholders. Governance mechanisms such as measurement of pre-defined metrics can help reduce the agency problem in one of two ways: by increasing goal alignment between the principal and the agent, or by reducing information asymmetry, e.g. making the agent's actions transparent to the principal. Agency theory suggests that certain governance mechanisms and contractual arrangements help increase goal alignment between a principal and an agent. In addition, information systems and task characteristics are proposed to moderate the effect of different governance arrangements on the agency risks [13].

Governance mechanisms targeting goal alignment include compensation contracts that tie agent pay to goal attainment by the principal, e.g., stock options for executives or bonuses for sales representatives. Governance mechanisms targeting information asymmetry include a means for limiting the agent's actions to those approved by the principal through the process of formalization and automation. This also helps increase the transparency of the agent's actions through management reporting, audit trails, and metric measurement.

2.3 BPM and the Process View of the Firm

In general, BPM can be defined as an organizational paradigm in which organizations are viewed as a collection of processes and managerial challenges. These are addressed through defining, analyzing, implementing, and continuously improving such processes. Within BPM, a process is defined as a network of activities performed in parallel or in sequence to achieve a desired outcome. In structured processes, the flow of activities, their inputs and outputs are well defined, described in organization-al policies and procedures, and enforced through workflow automation systems. However, well-structured processes are easily imitable and thus cannot be a source for competitive advantage [21], [22]. Unstructured processes, such as product development or strategic planning, are characterized by variability in the flow of

activities, expected outputs and available inputs. Moreover, the nature of work as well as re-sources necessary for accomplishing it are not well understood or purposefully kept ambiguous. Instead, organizations rely on process participants to search for available (and relevant) inputs and devise a way for converting such inputs into the best possible output. Unstructured processes are not easy to imitate and thus represent a more reliable source of competitive advantage.

BPM practitioners distinguish between transactional, development, enabling, and governing processes [4]. Transactional processes such as production, purchasing, fulfillment, and payroll are highly structured and support everyday business operations. Development processes such as product development and marketing are less structured and often generate information that is referenced by transactional processes, including product descriptions and prices, marketing and recruitment materials, etc. Development of human and IT resources is achieved through enabling processes, whereas governing processes are concerned with strategic planning, as well as risk and performance management [4].

Although activities are the key building blocks of processes, BPM is generally less concerned with how activities are performed. Rather, it focusses on what the inputs and outputs of such activities are, when the activities are performed (the flow), and to some degree, by whom they are performed (the actors) [23]. Activities receive inputs from other activities within or outside the process and convert them into output. Such outputs are used as inputs into other activities inside or outside of the process. Activities consume resources, which may include labor, information, or physical resource. However, the resources and their types are not well defined within the BPM frame-work and the issue of resource consumption is not fully addressed in BPM research [24].

In BPM, actor/roles include human process participants, computer information systems and potentially organizational units. Actors typically perform multiple activities and may be involved in several processes, although a certain level of specialization among actors is expected. Actors' time is viewed as a resource, and an actor cannot typically perform two activities at the same time [24]. In structured processes, actors have limited autonomy as their actions are either completely predefined (as is the case with code-based software or automated production processes), or severely constrained by workflow management systems. Unstructured processes are characterized by high actor autonomy and high agency risk. Therefore, actors' compensation plans are devised to ensure that they act in the best interests of the organization [18].

3 Conceptual framework

3.1 Agency in AI

The concept of agency is central to the field of Artificial Intelligence. In AI, an agent is defined as "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" [6]. An agent is rational if it selects actions that optimize its performance measure given available information. Information available to an agent comes from two sources, the prior knowledge provided to the agent by its designer and the information received by the agent through its percepts. A performance measure represents the key mechanism through which the goals of the agent are defined in relation to desirability of environmental states that result from the agent's actions.

An agent is considered autonomous to the extent it can compensate for partial or incorrect prior knowledge by learning from its actions and the percepts received from the environment. Therefore, ability to learn from its actions and from the data provided by its environment is considered a critical part of AI capability. Learning can be applied to different components of an agent, including the ability to infer relevant properties of the environment from percepts, resulting in the agent's own possible actions as well as the utility of information describing the desirability of world states [6, p. 694]. Consequently, machine learning is considered a key component of AI research and practice [1].

However, the ability to learn and the autonomy of an agent is restricted by the variety and format of percepts it can receive from its environment. The agent's autonomy is restricted to the extent to which it relies on human actors or human-designed processes as sources of information. As the agent gains the ability to accept and learn from stimuli directly from the environment through sensors or by means of computer vision or natural language processing, its autonomy increases.

3.2 AI and the Agency Problem

As in other types of principal agent relationships, agency risks in organizational use of AI stems from two sources, differences between the principal's and the agent's goals, and information asymmetries stemming from the lack of transparency about the agent's operations. Therefore, in order to understand the AI agency risks, it is important to define what the factors are that influence the level of alignment between the intelligent agent's goals vs. the principal's goals, and the level of transparency of AI operations.

Regarding AI and Goal Alignment, AI researchers and practitioners generally believe that it is possible and even necessary to assume that AI artifacts have goals. Significant research in the AI field focuses specifically on goal setting in AI [14], [15]. A rational AI agent will act to achieve environmental states that optimize its utility function. However, the agent's specific goals, at any point in time, are also influenced by its knowledge about the contribution of different environmental states to its utility function [6]. It is presumed that the ultimate utility function is defined by the designer of the AI; however, the contribution of different environmental states to the utility function may be a part of the a priori knowledge provided by the agent or by the designer or learned through reinforcement learning. In addition, agents typically have incomplete knowledge of environmental states, and therefore its goals will be influenced by its ability to infer the state of the environment from its percepts, the use of artificial inference. This ability is usually acquired by an agent through the process of supervised, semi-supervised, and unsupervised learning.

Assuming that the principal is an organization that deploys an AI artifact, its goal can be expected to optimize value for its stakeholders. Such a general goal is usually translated into a series of more detailed objectives that guide the performance of individual organizational units or processes. AI artifacts can be deployed within a particular business unit or business process. Therefore, for the purpose of this discussion, we will assume the principal's goals to be the objectives of the business unit or business process within which an AI artifact is deployed. We will consider separate situations when the same AI artifact is deployed in two different processes that have different objectives.

Let us examine the goal setting process for an AI artifact (the agent) in relation to the goals of a focal business unit (the principal). It is logical to assume that if an AI artifact is developed as a means to support the focal process, its utility function will be aligned with the process objectives. For example, a robot developed specifically for supporting a specific production process is expected to help optimize the objectives of that process. Similarly, a predictive model developed by an investment bank for identifying stock trading opportunities is expected to maximize the objectives of the process for which it is deployed, and presumably for the benefits of the bank as a whole [16], [17]. It is important to understand that to the extent to which an AI artifact is capable of learning from the environment, it is possible that its goals are distorted by exposure to biased environmental stimuli. For example, a trading algorithm trained on data from value stocks only is likely to underperform when asked to trade growth stocks or international securities. This introduces goal volatility of AI artifacts that are based on online and reinforcement learning algorithms.

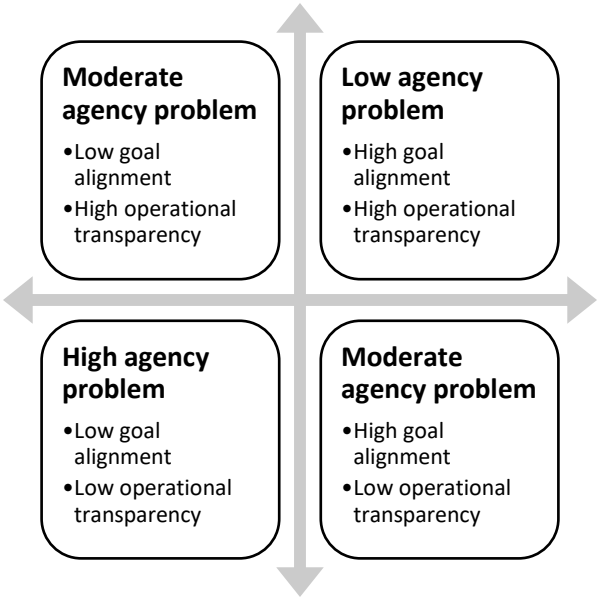
It is also possible that an AI artifact is developed by a third party and is deployed by a focal unit or process. Industry reports suggest that investments in AI are concentrated among a group of large technology companies and start-ups specializing in AI [1]. Such companies have access to the top AI talent and technology expertise. In addition, many such companies also have access to large volumes of data they collect as a by-product of providing their services; along with the purchase of data from data append providers such as Experian, Acxiom, and the Weather Company. They leverage such data for AI training [18], [19]. Therefore, it is likely that a large share of AI artifacts will be developed by these AI leaders, and deployed by other organizations in their business processes [20]. In these cases, it is logical to expect that the utility function of AI artifacts will be set up in such a way as to maximize the value for the company.

To the extent to which the AI developer is deriving value from the sale of the AI artifact, such artifacts will seek to maximize the value for its users. For example, a corporation selling large-scale AI applications to its corporate clients is expected to ensure that such applications deliver value for such clients. This is usually achieved by re-training a pre-trained model with tailored client data. This approach is referred to as transfer learning [21]. Organizational research on the agency suggests that the efforts of the AI developer to customize the AI artifact to the needs of the deploying company will be influenced by the presence of competition from other AI providers, outcome controls on the part of the client, and other factors.

On the other hand, a seller of consumer-focused AI artifacts, such as Amazon Echo or Google Home, is likely to configure the utility function of such devices in a way that maximizes the value of the provider as well as the consumer. One common way in which AI providers derive value is by collecting data from AI users to further train the AI platform [18]. One could expect that once such consumer data becomes less valuable [19], other AI providers will seek other sources of value such as using their AI to promote third party products and services to consumers. Therefore, one may expect goal alignment to be the lowest in the case of consumer or off-the-shelf AI.

In summary, the level of goal alignment between the principal (an organizational process) and the agent (an AI artifact) is expected to vary depending on a variety of factors. These include whether or not an AI artifact was built by the deploying organization or by a third party, along with the nature of contractual arrangements between the AI provider and the deploying organization. It will also be influenced by whether the artifact was

developed for the specific process or for a generic class of problems, and by the data used for the training of the artifact (see Figure 1).



3.2.1. Transparency in AI operations. Concerns about the lack of transparency of AI operations lie at the center of the discourse about the threat of AI [22]–[24]. The lack of transparency can be attributed to three factors: (1) computational algorithms and models that are difficult to understand/interpret especially as they become less stable and more adaptable, (2) lack of transparency regarding data sources for AI, especially as AI becomes capable of receiving data directly from the environment, and (3) the sheer speed and computational capacity of AI artifacts make their operations virtually impossible to audit.

Traditional software algorithms are designed to implement the rules devised by humans. In contrast, model-based software is expected to derive its rules from the data it is given. Conceptually, data driven rules are not new. Complex statistical models have long been the basis of many business and engineering applications. However, until recently, the models used in practice such as decision trees, linear and logistic regression, SVM models and case-based learning algorithms, have been amenable to human interpretation. Decision trees are the easiest to interpret in terms of business rules. Therefore, they have been heavily used in processes that require a high level of transparency due to regulatory considerations. Interpretation of regression models requires more extensive background knowledge, but is still rather straightforward. In comparison, advanced ML algorithms present a significant challenge in terms of interpretation. From ensemble models, to neural networks, the interpretability of the decision algorithms becomes more difficult. In addition, deep neural network models are nearly impossible to interpret, hence the creation of the Explainable AI (XAI) in the LIME project (Local Interpretable Model-Agnostic Explanations). Moreover, advanced AI is expected to rely on pipelines of models that are consecutively applied to data inputs, thus ensuring that such models represent a complete black box for its user. Complex AI models deliver superior predictive performance, and thus are expected to be adopted by organizations seeking to maximize return on investments. Yet, black box models and algorithms also significantly increase the risk of the agency problem.

The second key factor that leads to the decrease in the transparency of AI operations is the ability of AI to gather data directly from the environment, without human mediation. This ability is fostered by two trends. The first one is the proliferation of IoT devices which collect data through a series of sensors and share such data with other devices in the network [2]. As research in multi-agent negotiations advances, AI artifacts are expected to be able to broker agreements with other devices to access their data. Such agreements are likely to be numerous, expressed in computer code, and too costly for continuous human oversight. This will lead human actors to gradually relinquishing control over data access and sharing to AI. The second trend, continuous progress in representation learning, including image, video and speech recognition leads to increasing adoption of AI for data input into business processes, and for its use in pattern discovery in diverse unstructured data [21].

In summary, transparency of AI operations is influenced by several factors, including the learning and decision-making algorithms that are embedded into the AI artifacts, the complexity of the AI artifacts, the information gathering abilities of the artifacts, as well as by the system of explanations incorporated into the artifacts.

3.2.2. Analyzing AI agency risk. Combining the two factors, the risk for an agency problem increases as the levels of both operational transparency and goal alignment decrease. The agency problem is the lowest in situations where isolated ML solutions based on easily interpretable algorithms, such as decision trees or logistic regression, are developed in house for a specific process task. The agency problem increases as the AI solutions become more sophisticated, combine multiple deep learning algorithms and are able to receive data directly from the environment. Highly sophisticated AI algorithms pose an agency risk even when they are developed in-house. Such risks may stem from a poorly defined utility function or from AI learning undesirable

behavior from erroneous or biased data as referenced by Cathy O'Neil at a recent TED Talk on introducing human bias into objective algorithms (https://www.ted.com/speakers/cathy_o_neil). The agency problem increases as an organization outsources the development of AI artifacts to third parties thus relinquishing control over the utility function and the training data. To the extent that the sourced AI is relatively simple, the agency problem can be mitigated through the analysis of the underlying decision models, or through output control as in the case with simple image and speech recognition applications.

3.3 Applying AI agency framework to industry cases

Case 1. Anti-Money Laundering and Fraud is a significant challenge for many financial institutions, with losses amounting to millions of dollars every year. The industry standard software and methods in use today support the development of rules-based systems to catching fraudsters. Business users in financial institutions must think about what type of fraudulent transaction could take place or have happened in the past. They then create training data sets by identifying such types of events in historical data, and using analytics software to help develop business rules for identifying potential fraudulent transactions based on past events. These rules are then deployed to screening real-time transactions. Another approach would be to allow the machine to sift through the data and look for patterns in transactions in real time. The upside of this would be that relevant patterns not known before and not always detected by humans, can be identified and rules can be modified dynamically to account for such patterns. Notably, this would reduce AI operational transparency and heighten AI agency risk.

Case 2. Is it really you withdrawing money at the ATM? Deep Learning models executed on IBM's PowerAI platform were used in banks in China and the U.K. to take ATM camera data, analyze it in real time and determine if the user was covering up their face. If the algorithm inferred an issue, the ATM would immediately shut down and stop all transactions, along with contacting the authorities. ATM crimes of this nature have been significantly reduced. The case involves the use of a third-party AI platform. However, the agency risk is mitigated by the relative transparency of how the AI artifact is used in a business process. Such transparency allows banks to adjust the use of the algorithm for locations/seasons when partial face coverage is common, as in the middle of the winter in Siberia.

Case 3. A very large airport store with multiple locations used IoT video to analyze its customers while they were shopping. The ML/DL algorithms were trained on the images and classified customers into segments, such as a vacation vs. business customer. Based on their segment, specific marketing messages and selling routines were instituted to increase sales at each store. The pilot project was implemented in ten stores at an investment cost of $200K per store. The average increase in stores sales is $300K per year. They are now rolling this out to all airport store locations. Again, the relative transparency of the AI operation (it is used to segment customers into vacation and business) helps ensure that the AI is used in the best interests of the business owners.

3.4 Addressing AI agency risks through process management

Both application and development of AI are embedded in organizational process, and therefore, BPM methods can be used to identify, reduce or mitigate the AI agency risks. We propose that BPM approaches can contribute to addressing the AI agency risks through (1) explicit modeling of activities performed by AI, (2) explicit modeling of activities associated with development and training of AI, and (3) explicit modeling of the links between AI development/training and AI application activities.

3.4.1. Managing the use of AI in business processes. In spite of their resemblance of a system-based activity, an AI artifact is best conceptualized as an actor. In such capacity, it can be deployed by structured or unstructured processes. For example, image or speech recognition are routinely used in transactional processes including sales and payment processing or production. Increasingly, sophisticated AI is used in unstructured processes, from stock trading to R&D [1], [17]. In such processes, AI may be used as a tool for assisting human actors, or as an autonomous agent. However, the distinction between the two is rather diluted, as human actors may become complacent and delegate most decisions and responsibilities to AI artifacts.

The AI agency risks increase as more activities within a process are performed by AI and thus there is less opportunity for human actors to observe the outputs of such activities and exert outcome control. The agency theory suggests that in the absence of operational transparency, organizations are expected to impose stricter outcome controls on AI-heavy processes. However, this would require management to be aware of which processes rely on AI and to what extent. The lack of specialized notations for model-based or AI-based activities makes visualization of AI presence in a

process challenging. Another challenge comes from the fact that AI components are often incorporated into larger third-party software applications. Yet, as the number of such components increases, tracing the use of AI is the critical step in mitigating the agency risks presented by AI.

Based on the above discussion, we propose the following:

Proposition 1A: Explicit modeling of AI components in process maps is associated with a more accurate assessment of AI agency risks.

Proposition 1B: Explicit modeling of AI components in process maps is associated with adoption of more strict AI governance mechanisms.

3.4.2. Managing AI development and training. Like other actors, human or technology based, AI artifacts evolve over time. Within organizations, the evolution of actors is managed through dedicated business processes, including human resource development and IT development processes. Simple AI artifacts with limited learning capabilities are similar to IT and other technological assets that are designed and developed by humans. However, the design of AI artifacts is not based on the opinions of expert designers but on data. Therefore, in representing and managing AI development processes, special attention needs to be paid to training data, its sources and the governance procedures involved in data selection and validation. Therefore, it is recommended that processes focused on the development of AI and advanced analytics modeling be designed and managed separately from other IT development processes. Similarly, in the case of third party AI, AI acquisition processes may need to include activities not typically included in traditional IT acquisition processes, such as an audit of data sources used for AI training and AI retraining using evolving organizational data.

As AI artifacts develop an increased capacity for continuous and independent learning, AI development and acquisition processes become similar to those used for the management of human resources. Although organizations invest significant resources in training, such training is only partially responsible for the knowledge and skills possessed by their employees. Depending on the position, employees are hired and compensated for their knowledge and skills acquired through their formal and informal education, as well their work experience. Furthermore, employees learn on their jobs and potentially become more valuable to their future employers. In a similar manner, sophisticated AI artifacts are expected to come with pre-existing skills, be trained for a specific job, and be able to learn on the job. Depending on the AI ownership, AI artifacts may also be able to switch employers (in case of third party AI), and their value is expected to increase with their experience. Like employees, AI artifacts are capable of transferring knowledge within and outside of the organization. Therefore, over time, AI development processes need to include such activities as “resume” checking, onboarding, and making arrangements for non-disclosure of information. Special attention in such processes should be devoted to continuous efforts to maintain goal alignment through performance measurement and intentional retraining.

Based on the above discussion, we propose the following:

Proposition 2A: Organizations with specialized AI development and governance processes are more successful in accurately assessing AI agency risk than organizations that rely on traditional IT development and governance processes.

Proposition 2B: Organizations with specialized AI development and governance processes are more successful in mitigating AI agency risk than organizations that rely on traditional IT development and governance processes.

3.4.3. Managing autonomous AI learning. In a traditional approach to business analytics and AI development, AI training is a part of development processes and the trained AI artifact is deployed in a transactional or development process. Data from such processes is then fed back to the development process for future retraining and re-deployment of the AI artifact. In such situations, attention needs to be paid to accurately representing the linkages between development and deployment processes, especially as multiple AI artifacts may be involved and their re-training needs to be coordinated. As AI’s ability for independent learning grows, it becomes increasingly critical to trace the environments to which an AI artifact is exposed, and to be able to revert to a pre-exposure state of an artifact if the exposure results in being detrimental to the performance of the artifact.

Based on the above discussion, we propose the following:

Proposition 3A: Ability to trace AI artifact learning through exposure to training data and training environments is associated with a more accurate assessment of AI agency risks.

Proposition 3B: Ability to trace AI artifact learning through exposure to training data and training environments is associated with adoption of stricter AI and data governance mechanisms.

4 Discussion

The growth in AI capabilities and their increased deployment in organizational business processes presents opportunities for increased process efficiency and effectiveness. It also creates AI agency risks that need to be mitigated through appropriate AI governance mechanisms.

First, there is a need for theoretical and empirical research of organizational policies as they relate to AI use and governance, as well as on organizational adoption of AI governance mechanisms. While some of the AI governance policies may apply to all organizational processes, some other policies are expected to apply to some, but not other processes. Regulatory environment is expected to have a significant impact on adoption of AI governance policies. However, understanding how inherent process characteristics are related to AI use and AI governance for that process is a fruitful direction for IS research.

Second, there is a need for the development of process modeling standards that would support AI governance. Considering the critical role of data in AI development and training, it is important to represent which process data is used for AI training. It addition, increasing transparency of AI operations calls for explicit notations for activities that rely on machine learning and AI. Finally, there is a need for developing standard processes of AI development, acquisition, and governance that would take into account the unique aspects of AI. IS researchers and practitioners can play an important role in developing and evaluating such standards.

5 Conclusion

Increase in the autonomy of AI-enabled systems creates AI agency risks. In this paper, we applied agency theory and proposed a framework for the analysis of AI agency risks. The framework suggests that AI agency risks increase as the transparency of AI operations decreases and as the goal alignment between AI and the AI agent decreases (as is the case with third party AI). We further examined how BPM concepts can be applied to analyzing and mitigating AI agency problem.

The paper contributes to the IS research as it highlights the risks associated with AI assimilation within organizational processes and points towards fruitful directions for future research. Such directions include research on AI use and governance in business processes, development of process modeling standards for dealing with AI-enabled activities, and development of industry best practices for AI development, acquisition and governance processes.

6 References

[1] McKinsey Global Institute, “Artificial Intelligence: The new digital frontier?,” 2017.

[2] Gartner, “Gartner Says Artificial Intelligence Is a Game Changer for Personal Devices,” 2018. [Online]. Available: https://www.gartner.com/newsroom/id/3843263. [Accessed: 02-Mar-2018].

[3] Rage Frameworks, “The roald to Enterprise AI,” 2017. [Online]. Available: https://www.gartner.com/imagesrv/media-products/pdf/rage_frameworks/rage-frameworks-1-34JHQ0K.pdf.

[4] J. McCarthy, M. Minsky, N. Rochester, and C. E. Shannon, “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” 1955.

[5] A. M. Turing, “Computing Machinery and Intelligence,” *Mind. A Q. Rev. Psychol. Philos.*, vol. LIX, no. 236, pp. 433–460, 1950.

[6] S. Russell and peter Norvig, *Artificial Intelligence: A Modern Approach, 3rd Edition | Pearson*, 3rd ed. Pearson, 2010.

[7] E. Brynjolfsson and A. McAfee, *The second machine age : work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton and Company, 2014.

[8] S. A. Ross, “The Economic Theory of Agency: The Principal’s Problem,” *The American Economic Review*, vol. 63. American Economic Association, pp. 134–139, 1976.

[9] H. L. Tosi, L. R. Gomez-Mejia, and L. R. Gomez-Mejia, “The Decoupling of CEO Pay and Performance: An Agency Theory Perspective,” *Adm. Sci. Q.*, vol. 34, no. 2, p. 169, Jun. 1989.

[10] K. Roth and S. O’Donnell, “Foreign Subsidiary Compensation Strategy: An Agency Theory Perspective.,” *Acad. Manag. J.*, vol. 39, no. 3, pp. 678–703, Jun. 1996.

[11] K. M. Eisenhardt, “Agency and Institutional Theory Explanations: The Case of Retail Sales Compensation.,” *Acad. Manag. J.*, vol. 31, no. 3, pp. 488–511, Sep. 1988.

[12] L. Donaldson and J. H. Davis, “Stewardship Theory or Agency Theory: CEO Governance and Shareholder Returns,” *Aust. J. Manag.*, vol. 16, no. 1, pp. 49–64, Jun. 1991.

[13] K. M. Eisenhardt, “Agency Theory : An Assessment and Review,” *Acad. Manag. Rev.*, vol. 14, no. 1, pp. 57–74, 1989.

[14] J. O. Kephart and W. E. Walsh, “An artificial intelligence perspective on autonomic computing policies,” in *Proceedings. Fifth IEEE International Workshop on Policies for Distributed Systems and Networks*, 2004, pp. 3–12.

[15] W. E. Walsh, G. Tesauro, J. O. Kephart, and R. Das, “Utility functions in autonomic systems,” in *International Conference on Autonomic Computing, 2004. Proceedings.*, pp. 70–77.

[16] K. Maney, “Goldman Sacked: How Artificial Intelligence Will Transform Wall Street,”

Newsweek, 2017. [Online]. Available: http://www.newsweek.com/2017/03/10/how-artificial-intelligence-transform-wall-street-560637.html. [Accessed: 09-Mar-2018].

[17] K. Sennaar, "AI in Banking - An Analysis of America's 7 Top Banks," *Techemergence*, 2017. [Online]. Available: https://www.techemergence.com/ai-in-banking-analysis/. [Accessed: 09-Mar-2018].

[18] T. Moyniham, "Amazon Echo and Google Home Record What You Say. What Happens to Your Data? | WIRED," *Wired*, 2016. [Online]. Available: https://www.wired.com/2016/12/alexa-and-google-record-your-voice/. [Accessed: 09-Mar-2018].

[19] A. Agrawal, J. Gans, and A. Goldfarb, "Data May Be the New Oil, but Having Lots of It May Not Make You Rich," *Harvard Business Review*, 2018. [Online]. Available: https://hbr.org/2018/01/is-your-companys-data-actually-valuable-in-the-ai-era. [Accessed: 09-Mar-2018].

[20] M. Janakiram, "The Rise Of Artificial Intelligence As A Service In The Public Cloud," *Forbes*, 2018. [Online]. Available: https://www.forbes.com/sites/janakirammsv/2018/02/22/the-rise-of-artificial-intelligence-as-a-service-in-the-public-cloud/#67f8d579198e. [Accessed: 09-Mar-2018].

[21] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*, 1st ed. Boston: O'Reilly, 2017.

[22] N. Bostrom, "The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents," *Minds Mach.*, vol. 22, no. 2, pp. 71–85, May 2012.

[23] N. Bostrom, "Strategic Implications of Openness in AI Development 1."

[24] M. U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," *Harv. J. Law Technol.*, vol. 29, 2015.